OM nucleic - nucleic search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: September 23, 2000, 08:48:24 ; Search time 378.27 Seconds
(without alignments)
94.582 Million cell updates/sec

Title: US-09-428-674-11
Perfect score: 143
Sequence: 1 gtggpcatgtacttggctta........aattggttattaattgtt 143

Scoring table: OLIGO_NUC
Searched: Gapop 60.0 , Gapext 60.0
                311585 seqs, 125096042 residues
Word size :   15

Total number of hits satisfying chosen parameters: 88

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Match Length	DB	ID	Description
c 1	21	14.7	84	1	T21573	Human gene signatu
c 2	17	11.9	4121	1	V74626	Staphylococcus aur
c 3	17	11.9	5790	1	X24464	Bovine myostatin g
c 4	17	11.9	6029	1	X23824	Bacillus sp. GUS g
c 5	16	11.2	200	1	Q77245	Human genome fragm
c 6	16	11.2	211	1	T19250	Human gene signatu
c 7	16	11.2	226	1	T21161	Human gene signatu
c 8	16	11.2	692	1	V69554	Human gene signatu
c 9	16	11.2	800	1	X39802	U. florida type I
c 10	16	11.2	1689	1	066824	Gastric cancer ass
c 11	16	11.2	2132	1	066392	AmEPV NPH-1 gene.
c 12	16	11.2	2834	1	161896	B.thuringiensis 1
c 13	16	11.2	3073	1	014635	Human 14-3-3 zeta
c 14	16	11.2	3073	1	T34371	Plasmid pATG29 (AT
c 15	16	11.2	8457	1	066797	AmEPV spheroidin g
c 16	16	11.2	110000	1	V21209_00	Methanococcus jann
c 17	16	11.2	110000	1	X20248_08	Continuation (9 of
c 18	16	11.2	132	1	T19753	Human gene signatu
c 19	16	11.2	192	1	085552	DNA probe 58 detec
c 20	15	10.5	326	1	060587	Human brain Expres
c 21	15	10.5	343	1	T25391	Human gene signatu
c 22	15	10.5	607	1	V75299	Staphylococcus aur
c 23	15	10.5	750	1	V53475	DNA encoding a Sta
c 24	15	10.5	885	1	N20095	Sequence of leukoc
c 25	15	10.5	885	1	N20010	Hybrid human leuko
c 26	15	10.5	888	1	046120	E. coli 16S rRNA f
c 27	15	10.5	937	1	N40112	Sequence of the cD
c 28	15	10.5	937	1	N30105	Sequence on recomb
c 30	15	10.5	961	1	N40107	Sequence on recomb
c 31	15	10.5	961	1	N30100	Sequence on recomb
c 32	15	10.5	1177	1	X08427	LapA2 promoter seq
c 33	15	10.5	1298	1	V32590	Schwannomin-bindin
	15	10.5	1498	1	013005	16S rRNA gene corr
c 34	15	10.5	1540	1	N91514	Escherichia coli 1
c 35	15	10.5	1542	1	T18759	E. coli 16S riboso
c 36	15	10.5	1542	1	T29140	E. coli 16S rRNA (
c 37	15	10.5	1542	1	Q49060	rRNA gene (irsE) f
c 38	15	10.5	1743	1	T85939	16S rRNA gene of S
c 39	15	10.5	1743	1	V53502	Shigella flexneri
c 40	15	10.5	1866	1	N90787	DNA encoding a Sta
c 41	15	10.5	2218	1	T40700	Clone L19 of Brass
c 42	15	10.5	2224	1	N90787	N. benthamiana phy
c 43	15	10.5	2273	1	Q04339	Aquaricin I gene
c 44	15	10.5	2274	1	Q12818	Aquaricin I coding
c 45	15	10.5	2274	1	Q75859	Aqualysin I gene.

ALIGNMENTS

RESULT 1
ID   T21573
T21573 standard; cDNA to mRNA; 84 BP.
AC   T21573;
DT   03-AUG-1996 (first entry)
DE   Human gene signature HUMGS02953.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; JO1916.
PR   12-NOV-1993; JP-355504.
PA   (OKUB/) OKUBO K.
PA   (MATS/) MATSUBARA K.
PI   Matsubara K, Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 916; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence 84 BP; 28 A; 16 C; 19 G; 21 T;

Query Match   14.7%; Score 21; DB 1; Length 84;
Best Local Similarity 100.0%;
Matches   21; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy 112 aaaaattaataaattgggttt 132
         |||||||||||||||||||||
Db   53 AAAAATTAATAAATTGGGTTT 73

RESULT 2
ID   V74626
V74626 standard; DNA; 4121 BP.
AC   V74626;
DT   26-MAR-1999 (first entry)
DE   Staphylococcus aureus contig SEQ ID #315.

```
repeat_region    16352..16665
                 /rpt_family="Alu"
repeat_region    16666..17036
                 /rpt_family="Alu"
repeat_region    17037..17102
                 /rpt_family="MaLR"
repeat_region    17103..17398
                 /rpt_family="L1"
repeat_region    17399..17739
                 /rpt_family="Alu"
repeat_region    17740..17889
                 /rpt_family="L1"
repeat_region    17890..18198
                 /rpt_family="Alu"
repeat_region    18199..18335
                 /rpt_family="L1"
repeat_region    18336..18441
                 /rpt_family="Alu"
repeat_region    18446..18459
                 /rpt_family="Alu"
repeat_region    18460..18482
                 /rpt_family="Alu"
repeat_region    18483..18725
                 /rpt_family="(TAAA)n"
repeat_region    18725..?
                 /rpt_family="L2"
repeat_region    19043..19335
                 /rpt_family="AT_rich"
repeat_region    19358..19511
                 /rpt_family="Alu"
repeat_region    20026..20068
                 /rpt_family="MIR"
repeat_region    20129..20174
                 /rpt_family="L2"
repeat_region    20218..20529
                 /rpt_family="AT_rich"
repeat_region    20702..20764
                 /rpt_family="AT_rich"
misc_feature     20991..21066
                 /note="match to EST R85763 (NID:g944169) yq23c06.s1"
repeat_region    21235..21527
                 /rpt_family="Alu"
repeat_region    21997..22102
                 /rpt_family="L2"
repeat_region    23226..23370
                 /rpt_family="MIR"
repeat_region    23461..23932
                 /rpt_family="MIR"
misc_feature     23656..23913
                 /note="match to EST AA526665 (NID:g2264593) n168b12.s1"
misc_feature     23702..23890
                 /note="match to EST AA228421 (NID:g1849952) nc39b09.r1"
misc_feature     23702..23890
                 /note="similar to Mus musculus EST AU035453"
misc_feature     ..
                 /note="similar to Mus musculus EST AI119247 (NID:g3519571)"
```

```
RESULT 10
AL136180

LOCUS       AL136180    99487 bp    DNA    HTG    01-FEB-2000
DEFINITION  Homo sapiens chromosome 6 clone RP3-477I23, *** SEQUENCING IN
            PROGRESS ***, 8 unordered pieces.
ACCESSION   AL136180
VERSION     AL136180.1  GI:6982309
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
```

```
Query Match     21.4%;   Score 45;  DB 39;  Length 83928;  Pred. No. 1.2e-15;
Best Local Similarity  100.0%;
Matches 45;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy    116 aagtgatctcccacctaggcctcccaaagtgctgggattacagg 160
          ||||||||||||||||||||||||||||||||||||||||||||
Db  75844 AAGTGATCTCCCACCTAGGCCTCCCAAAGTGCTGGGATTACAGG 75800
```

```
COMMENT     On Feb 16, 2000 this sequence version replaced gi:6689799.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in
            progress and the release of this data is based on the understanding
            that the sequence may change as work continues. The sequence may
            be contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Contig ID: 00017 Length: 1162bp
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *          1  1162: contig of 1162 bp in length
            *       1163  1162: gap of 800 bp in length
            *       1963  1962: contig of 13765 bp in length
            * Contig_ID: 00227 Length: 13765bp
            *      15728 15727: gap of 800 bp in length
            *      16528 16527: contig of 1384 bp in length
            * Contig_ID: 00138 Length: 1384bp
            *      17912 17911: gap of 800 bp in length
            *      18712 18711: contig of 42612 bp in length
            * Contig_ID: 00709 Length: 42612bp
            *      61324 61323: gap of 800 bp in length
            *      62124 62123: contig of 1082 bp in length
            * Contig_ID: 00828 Length: 1082bp
            *      63206 63205: gap of 800 bp in length
            *      64006 64005: contig of 800 bp in length
            * Contig_ID: 00896 Length: 800bp
            *      64806 64805: gap of 800 bp in length
            *      65156 65155: contig of 5150 bp in length
            * Contig_ID: 00906 Length: 5150bp
            *      69956 69955: gap of 800 bp in length
            *      71423 71422: contig of 1467 bp in length
            * Contig_ID: 00963 Length: 1467bp
            *      72223 72222: gap of 800 bp in length
            *      99487: contig of 27265 bp in length
            * Contig_ID: 00963 Length: 27265bp.
FEATURES             Location/Qualifiers
     source          1..99487
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP3-477B23"
                     /clone_lib="RPCI-3"
BASE COUNT    26800 a  18017 c  17863 g  31206 t  5601 others
ORIGIN
```

```
RESULT 11
AC004977

LOCUS       AC004977   106474 bp    DNA    HTG    12-JUN-1998
DEFINITION  Homo sapiens clone DJ1152C17, *** SEQUENCING IN PROGRESS ***, 1
            unordered pieces.
ACCESSION   AC004977
VERSION     AC004977.1  GI:3213020
```

```
Query Match     21.4%;   Score 45;  DB 32;  Length 99487;  Pred. No. 1.2e-15;
Best Local Similarity  100.0%;
Matches 45;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy    126 cccacctaggcctcccaaagtgctgggattacaggcgtgagccac 170
          |||||||||||||||||||||||||||||||||||||||||||||
Db  72711 CCCACCTAGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCAC 72755
```

KEYWORDS   HTG; HTGS_PHASE1.
SOURCE   human.
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 106474)
AUTHORS   Waterston,R.H.
TITLE   The sequence of Homo sapiens clone
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 106474)
AUTHORS   Waterston,R.H.
TITLE   Direct Submission
JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
COMMENT   * NOTE: This is a 'working draft' sequence. It currently
              * consists of 1 contig. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              *
              * 1 106474: contig of 106474 bp in length.
FEATURES   Location/Qualifiers
     source   1..106474
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DJ1152C17"
BASE COUNT   28957 a 22910 c 23777 g 30830 t
ORIGIN

Qy   126 cccacctaagcctccaaagtgctggdattacaggcatggagccac 170
          ||||||||||||||||||||||||||||||||||||||||||||||
Db 73539 CCCACCTAAGCCTCCCAAAGTGCTGGGATTACAGGCATGGAGCCAC 73583

RESULT 12
HSJ635E18 HSJ635E18 117859 bp   DNA   HTG   20-APR-2000
LOCUS
DEFINITION   Homo sapiens chromosome 1 clone RP4-635E18 map p36.11-36.31, ***
SEQUENCING IN PROGRESS *** in unordered pieces.
ACCESSION   AL109811
VERSION   AL109811.21 GI:7635714
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   human.
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 117859)
AUTHORS   Walls,J.
TITLE   Direct Submission
JOURNAL   Submitted (20-APR-2000) Sanger Centre, Hixton, Cambridgeshire,
              CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
              requests: clonerequests@sanger.ac.uk
COMMENT   On Apr 22, 2000 this sequence version replaced gi:7242677.
              -------------- Genome Center
              Center: Sanger Centre
              Center code: SC
              Web site: http://www.sanger.ac.uk
              Contact: humquery@sanger.ac.uk
              -------------- Project Information
              Center project name: dJ635E18
              -------------- Summary Statistics
              Assembly program: XGAP4; Version 4.5
              Sequencing vector: M13; M77815; 14% of reads
              Sequencing vector: pUC6mid; L08752; 85% of reads
              Chemistry: Dye-terminator Abi; 1% of reads

Chemistry: Dye-terminator Big Dye; 94% of reads
Chemistry: Dye-primer-amersham; 4% of reads
Consensus quality: 112388 bases at least Q40
Consensus quality: 113891 bases at least Q30
Consensus quality: 114841 bases at least Q20
Insert size: 116859; sum-of-contigs
Insert size: 123727; 14.6% error; agarose-fp
Quality coverage: 5.57x in Q20 bases; sum-of-contigs Quality
coverage: 5.26x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently * consists
* of 11 contigs. The true order of the pieces is * not known and
* their order in this sequence record is arbitrary. Where the
* contigs adjacent to the vector can * be identified, they are
* labelled with 'clone_end' in the * feature table. Some order and
* orientation information * can tentatively be deduced from paired
* sequencing reads * which have been identified to span the gap
* between two * contigs. These are labelled as part of the same *
* 'fragment_chain', and the order and relative orientation * of the
* pieces within a fragment_chain is reflected in * this file. Gaps
* between the contigs are represented as * runs of N, but the exact
* sizes of the gaps are unknown. * This record will be updated with
* the finished sequence as * soon as it is available and the
* accession number will be * preserved.
*
* 1   55698: contig of 55698 bp in length
* 55799   72183: contig of 16385 bp in length; fragment_chain 1 *
* 72284   73838: contig of 1555 bp in length; fragment_chain 1 *
* 73939   75462: contig of 1524 bp in length
* 75563   76927: contig of 1365 bp in length
* 77028   78261: contig of 1234 bp in length
* 78362   79923: contig of 1562 bp in length
* 80024   81095: contig of 1072 bp in length
* 81196   112651: contig of 31456 bp in length; fragment_chain 2
* 112752 115113: contig of 2362 bp in length; fragment_chain 2 *
* 115214 117859: contig of 2646 bp in length; fragment_chain 1 *
FEATURES   Location/Qualifiers
     source   1..117859
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="p36.11-36.31"
                     /clone="RP4-635E18"
                     /clone_lib="RPCI-4"
     misc_feature   1..55698
                     /note="assembly_fragment:00832"
     misc_feature   55799..72180
                     /clone_end:SP6
                     /vector_side:left"
     misc_feature   55799..72284
                     /note="assembly_fragment:02428
                     /fragment_chain:1"
     misc_feature   72284..73838
                     /note="assembly_fragment:02464
                     /fragment_chain:1"
     misc_feature   73939..75462
                     /note="assembly_fragment:00023"
     misc_feature   75563..76927
                     /note="assembly_fragment:00489"
     misc_feature   77028..78261
                     /note="assembly_fragment:00638"
     misc_feature   78362..79923
                     /note="assembly_fragment:00755"
     misc_feature   80024..81095
                     /note="assembly_fragment:01032"
     misc_feature   81196..112651
                     /note="assembly_fragment:02380
                     /fragment_chain:2"
     misc_feature   112752..115113
                     /note="assembly_fragment:02462
                     /fragment_chain:2"

Query Match   21.4%; Score 45; DB 41; Length 106474;
Best Local Similarity 100.0%; Pred. No. 1.2e-15;
Matches   45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
117:  gb_gss13:*
118:  gb_gss14:*
119:  gb_gss15:*
120:  gb_gss16:*
121:  gb_gss17:*
122:  gb_gss18:*
123:  gb_gss13:*
124:  em_gss13:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	74	16.3	441	22	AA157818	AA157818 zo35h07.s
c	2	71	15.7	417	21	AA099427	AA099427 zl79g10.s
c	3	71	15.7	431	20	AA052890	AA052890 zl71e08.s
c	4	71	15.7	439	21	AA135403	AA135403 zo28e05.s
c	5	71	15.7	447	20	AA055663	AA055663 z175f05.s
c	6	70	15.5	333	27	AA516374	AA516374 nf59g03.s
c	7	70	15.5	435	21	AA135390	AA135390 zo28b04.s
c	8	69	15.2	501	62	AW007800	AW007800 wt03c11.x
c	9	69	15.2	333	93	AQ045412	AQ045412 RPCI11-32
c	10	65	14.3	333	93	AI092628	AI092628 qa35c12.x
c	11	65	14.3	431	35	AL134462	AL134462 DKFZp547J
c	12	65	14.3	643	47	AA991933	AA991933 os9b03.s
c	13	63	13.9	136	33	AI584112	AI584112 ts13h03.x
c	14	62	13.7	401	41	AA034378	AA034378 zk20c10.s
c	15	61	13.5	439	20	AA419145	AA419145 zv34h08.s
c	16	59	13.5	176	26	AA807651	AA807651 nx08c07.s
c	17	59	13.0	289	31	AA243224	AA243224 zr26g06.s
c	18	59	13.0	292	23	AA224589	AA224589 zr17e10.s
c	19	59	13.0	331	23	AA053427	AA053427 zl71h06.s
c	20	59	13.0	333	20	AA219559	AA219559 zq99c04.s
c	21	59	13.0	339	22	AA216639	AA216639 zq95f07.s
c	22	58	13.0	569	22	AA250803	AA250803 zs06f12.s
c	23	58	12.8	184	23	AA832265	AA832265 oc91f07.s
c	24	58	12.8	298	31	AA418333	AA418333 zv96c05.r
c	25	58	12.8	420	26	AI242047	AI242047 qh81c06.x
c	26	58	12.8	437	37	W02703	W02703 zc64d08.s1
c	27	57	12.8	413	90	AI264351	AI264351 qi09e06.x
c	28	57	12.6	553	37	AA424412	AA424412 zv82g12.s
c	29	56	12.6	298	26	AW665275	AW665275 hj02a08.x
c	30	56	12.4	426	79	AA478483	AA478483 zx6d08.s1
c	31	55	12.4	144	27	AQ238587	AQ238587 RPCI11-63
c	32	55	12.1	379	95	R67329	R67329 yi32d03.s1
c	33	55	12.1	405	88	R77223	R77223 yi17b12.s1
c	34	55	12.1	423	89	AI274932	AI274932 qi49d04.x
c	35	55	12.1	430	37	T58256	T58256 yb50d07.s1
c	36	55	12.1	445	90	AA975725	AA975725 oq64a11.s
c	37	55	12.1	513	33	AA479719	AA479719 zu43h04.s
c	38	55	12.1	527	27	AQ586372	AQ586372 RPCI-11-4
c	39	54	12.1	589	106	T93522	T93522 ye14e05.s1
c	40	54	11.9	378	85	H03663	H03663 yj37g11.s1
c	41	54	11.9	379	90	T93545	T93545 ye14e05.s1
c	42	54	11.9	380	79	AW630829	AW630829 hh88b05.y
c	43	54	11.9	488	95	AQ240648	AQ240648 CIT-HSP-2
c	44	54	11.9	601	96	AQ309497	AQ309497 CIT-HSP-2
c	45	53	11.7	191	22	AA204919	AA204919 zq62h07.s

ALIGNMENTS

```
RESULT 1
AA157818
AA157818/c
LOCUS       AA157818    441 bp    mRNA    EST    16-DEC-1996
DEFINITION  zo35h07.s1 Stratagene colon (#937204) Homo sapiens cDNA clone
            IMAGE:588925 3' similar to contains LTR7.t3 LTR7 repetitive element
ACCESSION   AA157818
VERSION     AA157818.1  GI:1732647
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 441)
AUTHORS     Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
TITLE       Generation and analysis of 280,000 human expressed sequence tags
JOURNAL     Genome Res. 6 (9), 807-828 (1996)
MEDLINE     97064478
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40M13 fwd. from Amersham
            High quality sequence stop: 297.
FEATURES             Location/Qualifiers
     source          1..441
                     /organism="Homo sapiens"
                     /db_xref="GDB:4626665"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:588925"
                     /clone_lib="Stratagene colon (#937204)"
                     /tissue_type="tumor"
                     /cell_line="T84 carcinoma cell line"
                     /lab_host="SOLR cells (kanamycin resistant)"
                     /note="Organ: colon; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
                     EcoRI; Oligo dT-84 colonic epithelial cell line. Average
                     Insert size: 1.0 kb; Un1-ZAP XR Vector; -5' adaptor
                     sequence: 5' GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
                     CTCGAGTTTTTTTTTTTTTTTTTT 3'"
BASE COUNT      116 a     75 c    129 g    119 t      2 others
ORIGIN

Query Match  16.3%;   Score 74;  DB 22;  Length 441;
Best Local Similarity  100.0%;   Pred. No. 1.7e-29;
Matches  74;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   301  cccttgctgactcctcttttcggactcagccccgccaggtgaaataaaacagcct  360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    74  CCCTTTGCTGACTCTCTTTTTTCGGACTCAGCCCGCCGCCTGCACCCCAGGTGAAATAAAACAGCCT  15

Qy   361  tgttgctcacaaa  374
          |||||||||||||
Db    14  TGTTGCTCACAAAA  1


RESULT 2
AA099427
AA099427/c
LOCUS       AA099427    417 bp    mRNA    EST    21-JUL-1997
DEFINITION  zl79g10.s1 Stratagene colon (#937204) Homo sapiens cDNA clone
            IMAGE:510882 3' similar to contains LTR7.t3 LTR7 repetitive element
ACCESSION   AA099427
VERSION     AA099427.1  GI:1646199
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
```

COMMENT   On Jan 6, 2000 this sequence version replaced gi:6676376.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: 40UP from Gibco.

FEATURES   Location/Qualifiers
     source   1..399
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /clone="IMAGE:2909872"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo."

BASE COUNT   106 a   87 c   87 g   119 t
ORIGIN

RESULT 3
AI076014/c
LOCUS   AI076014   407 bp   mRNA   EST   27-AUG-1998
DEFINITION ov47d10.x1 Soares_testis_NHT Homo saapiens cDNA clone IMAGE:1640467
            3', mRNA sequence.
ACCESSION   AI076014
VERSION   AI076014.1 GI:3405192
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 407)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   On Apr 7, 1998 this sequence version replaced gi:3036064.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www.bio.llnl.gov/bbrp/image/image.html
            Insert Length: 480   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham.

Query Match   23.8%; Score 82; DB 72; Length 399;
Best Local Similarity 100.0%; Pred. No. 1.5e-28;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 tgcctccagagaagaacgcagccctactgacaccttgggttttggcgttaagaaccaact 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 394 TGCCTCCAGAGAAGAACGCAGCCCTACTGACACCTTGGGTTTTGGCCGTTAAGAACCAACT 335

Qy   61 tggacttttcacttccaaact 82
        |||||||||||||||||||||
Db 334 TGGACTTTTCACTTCCAAAACT 313

FEATURES   Location/Qualifiers
     source   1..407
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Soares_testis_NHT"
                     /clone="IMAGE:1640467"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech
                     Laboratories, Inc., and primed with a Not I - oligo(dT)
                     primer [5'...GCGGCCGCCCCCAATTTTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT   105 a   91 c   88 g   123 t
ORIGIN

RESULT 4
AI914390/c
LOCUS   AI914390   431 bp   mRNA   EST   17-DEC-1999
DEFINITION wd48d03.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2331365 3', mRNA sequence.
ACCESSION   AI914390
VERSION   AI914390.1 GI:5634245
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 431)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            Insert Length: 626   Std Error: 0.00
            High quality sequence stop: 430.

FEATURES   Location/Qualifiers
     source   1..431
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /clone="IMAGE:2331365"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed and ss circles were made in

Query Match   23.8%; Score 82; DB 35; Length 407;
Best Local Similarity 100.0%; Pred. No. 1.5e-28;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 tgcctccagagaagaacgcagccctactgacaccttgggttttggcgttaagaaccaact 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 403 TGCCTCCAGAGAAGAACGCAGCCCTACTGACACCTTGGGTTTTGGCCGTTAAGAACCAACT 344

Qy   61 tggacttttcacttccaaact 82
        |||||||||||||||||||||
Db 343 TGGACTTTTCACTTCCAAAACT 322

```
117:  gb_gss13:*
118:  gb_gss14:*
119:  gb_gss15:*
120:  gb_gss16:*
121:  gb_gss17:*
122:  gb_gss18:*
123:  gb_gss19:*
124:  em_gss13:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	21	4.4	213	85	H33155	H33155 EST108874 R
2	21	4.4	231	33	AA963360	AA963360 UI-R-E1-g
3	21	4.4	231	33	AA955831	AA955831 UI-R-E1-f
4	21	4.4	286	33	AA998842	AA998842 UI-R-C0-1
5	20	4.2	412	93	AQ055186	AQ055186 CIT-HSP-2
6	20	4.2	163	93	AI283995	AI283995 qt66e09.x
7	19	4.0	526	37	AI287912	AI287912 qu13c03.x
8	19	4.0	419	47	AL042275	AL042275 DKFZp434K
9	19	4.0	474	37	AI287912	AI287912 qu13c03.x
10	19	4.0	506	47	AL046807	AL046807 DKFZp434O
11	19	4.0	710	105	AQ482369	AQ482369 RPCI-11-2
12	18	3.8	239	91	W85317	W85317 mf53h12.r1
13	18	3.8	251	63	W85317	W85317 mf53h12.r1
14	18	3.8	300	93	AQ000680	AQ000680 CIT-HSP-2
15	18	3.8	303	30	AA702247	AA702247 z184g08.s
16	18	3.8	330	119	AZ061388	AZ061388 RPCI-23-4
17	18	3.8	352	93	AQ008470	AQ008470 CIT-HSP-2
18	18	3.8	361	95	AQ200054	AQ200054 RPCI11-44
19	18	3.8	367	40	AI467535	AI467535 ve35c06.x
20	18	3.8	376	102	AQ418263	AQ418263 RPCI-11-1
21	18	3.8	387	94	AQ094468	AQ094468 HS_3023/B
22	18	3.8	395	29	AA645356	AA645356 vs79g12.r
23	18	3.8	420	35	AI073153	AI073153 UI-R-Y0-m
24	18	3.8	427	110	AQ752975	AQ752975 HS_5226_B
25	18	3.8	429	120	B38668	B38668 HS_7048-A1-
26	18	3.8	446	105	AQ536009	AQ536009 RPCI-11-4
27	18	3.8	449	79	AW670115	AW670115 r13892 MA
28	18	3.8	458	102	AQ450037	AQ450037 500007C01
29	18	3.8	459	29	AA667436	AA667436 vv1g07.r
30	18	3.8	478	95	AQ239927	AQ239927 CIT-HSP-2
31	18	3.8	553	36	AI180247	AI180247 EST223990
32	18	3.8	557	79	AW631901	AW631901 91282 MAR
33	18	3.8	594	95	AQ234084	AQ234084 HS_2057_A
34	18	3.8	603	44	AJ274271	AJ274271 HS_2057_A
35	18	3.8	614	119	AZ044562	AZ044562 Gm_Umb001
36	18	3.8	619	122	Z89685	Z89685 F.rubripes
37	18	3.8	695	95	AQ194820	AQ194820 RPCI11-65
38	18	3.8	700	96	AQ309735	AQ309735 CIT-HSP-2
39	18	3.8	711	101	AQ330194	AQ330194 nbxb0046N
40	18	3.8	715	95	AQ240677	AQ240677 CIT-HSP-2
41	18	3.8	728	47	AJ274313	AJ274313 AJ274313
42	18	3.8	740	102	AQ449564	AQ449564 50000lpi1
43	18	3.8	779	35	AI096883	AI096883 qb58f02.x
44	17	3.8	809	90	W22729	W22729 75C5 Human
45	17	3.6	174	92	Z19745	Z19745 HSAAALEW H
	17	3.6	183	59	AV242584	AV242584 AV242584

ALIGNMENTS

RESULT 1

LOCUS   H33155   213 bp   mRNA   EST   02-APR-1998
DEFINITION EST108874 Rat PC-12 cells, NGF-treated (9 days) Rattus sp. cDNA
            clone RPNA121 5' end, mRNA sequence.
ACCESSION   H33155
VERSION   H33155.1 GI:978572
KEYWORDS   EST.
SOURCE   Rattus sp.
  ORGANISM Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 213)
  AUTHORS   Lee,N.H., Weinstock,K.G., Kirkness,E.F., Earle-Hughes,J.A.,
            Fuldner,R.A., Marmaros,S., Glodek,A., Gocayne,J.D., Adams,M.D.,
            Kerlavage,A.R., Fraser,C.M. and Venter,J.C.
  TITLE   Comparative expressed-sequence-tag analysis of differential gene
            expression profiles in pc-12 cells before and after nerve growth
            factor treatment
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 92, 8303-8307 (1995)
  MEDLINE   95396786
COMMENT   On Jan 6, 2000 this sequence version replaced gi:6676554.
            Other_ESTs: EST108873
            Contact: Lee, NH
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            For clone availability please contact the TIGR Database
            (tdbinfo@tdb.tigr.org)
            Seq primer: M13 Reverse.
FEATURES   Location/Qualifiers
     source   1..213
                     /organism="Rattus sp."
                     /db_xref="ATCC (inhost):2003416"
                     /db_xref="taxon:10118"
                     /clone_lib="Rat PC-12 cells, NGF-treated (9 days)"
                     /clone="RPNA121"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; poly(A)+ RNA was purified from 9-day NGF treated
                     PC12 cells; cDNA was constructed using an oligo-dT primer
                     and directionally cloned using the Lambda ZAP II Vector
                     Kit by Stratagene"
BASE COUNT   73 a   43 c   51 g   44 t   2 others
ORIGIN

Query Match   4.4%; Score 21; DB 85; Length 213;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 300 gagtgaggaagaagtgaga 320
        |||||||||||||||||||
Db   2 GAGTGAGGAAGAGAGTGAGA 22

RESULT 2

LOCUS   AA963360   231 bp   mRNA   EST   09-MAR-1999
DEFINITION UI-R-E1-ga-g-05-0-UI.s1 UI-R-E1 Rattus norvegicus cDNA clone
            UI-R-E1-ga-g-05-0-UI 3', mRNA sequence.
ACCESSION   AA963360
VERSION   AA963360.1 GI:4278284
KEYWORDS   EST.
SOURCE   Rattus norvegicus.
  ORGANISM Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 231)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE   Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044472
```